

University of Tartu
Institute of Philosophy and Semiotics

KANT ON MANIPULATION AND PERSONAL AUTONOMY

Master's Thesis in Philosophy

Semen Reshenin

Supervisor: Eva Piirimäe

2017

INTRODUCTION.....	3
1 ON THE POSSIBILITY OF A KANTIAN EMPIRICAL PSYCHOLOGY	4
2 A BRIEF ACCOUNT OF KANT'S EMPIRICAL PSYCHOLOGY	8
2.1 HUMAN ACTION	8
2.2 HUMAN COGNITION.....	12
2.3 DEFECTS OF COGNITION AND VOLITION.....	14
3. NATURE AS AN ORIGINAL MANIPULATOR.....	18
3.1 KANT'S NATURAL TELEOLOGY	19
3.2 THE UNSOCIAL SOCIABILITY	21
4 THE MECHANISM OF MANIPULATION.....	24
4.1 CONFORMITY.....	25
4.2 AMBITION AND OTHER PASSIONS	27
4.3 ENTHUSIASM AND OTHER AFFECTS	28
4.4 CHANGING PUBLIC OPINION.....	30
4.5 THE ATTRACTIVENESS OF CONSPIRACY THEORIES.....	31
4.6 THE CONCEPT OF MANIPULATION	32
5 PERSONAL AUTONOMY AS A SOLUTION.....	33
5.1 MORAL AUTONOMY AND PERSONAL AUTONOMY	35
5.2 WHO IS TO BLAME?.....	39
6 WHAT CAN WE DO?	40
6.1 THE MAXIM OF REASON'S SELF-PRESERVATION.....	41
6.2 THE MAXIM OF A BROAD-MINDED WAY OF THINKING	43
6.3 THE MAXIM OF CONSISTENCY	45
CONCLUSION	46
ABSTRACT	46
REFERENCES	47

Introduction

People as social beings regularly get each other to do what they want in a variety of ways. But not every way is morally unproblematic. According to Immanuel Kant, we are doing something wrong when we treat others merely as means, and not also as ends in themselves. We ought to respect others as rational beings, and manipulation is incompatible with such respect. At the same time, Kant believes that people should have courage to use their own understanding without the direction of another. If a person lacks such courage, then she lives in the state of self-incurred immaturity, which makes her particularly susceptible to manipulation. For Kant, her unwillingness to use her own understanding is a failure to respect humanity in her own person; hence it is a moral failure.

My project can be summed up as the reconstruction of Kant's theory of manipulation. While I touch upon the issues of the moral responsibility for manipulation and its political applications, my main focus is on the mechanism of manipulation. In my thesis I argue (1) that Kant's pragmatic anthropology and empirical psychology can provide us with an explanation why people are susceptible to manipulation; and (2) that Kant's maxims of common understanding can help us partially protect ourselves from the deteriorating effects of manipulation by establishing personal autonomy.

My thesis heavily relies on the contemporary discussions on the empirical dimension of Kant's moral philosophy. I owe much to Wood (1999; 2007), Louden (2000; 2011), Frierson (2005; 2014), who did a great job to show that Kant should not be regarded as psychologically naïve. I will show that a framework of Kant's empirical psychology is robust enough to explain how manipulation works and why it is so efficient. I believe that a Kantian theory of manipulation is relevant for contemporary discussions about manipulation because it has good explanatory power and provides us with an unorthodox perspective on a manipulatee. It considers a manipulatee as prudentially and morally responsible for being vulnerable in the face of manipulation. What should be stressed is that this kind of responsibility does not justify manipulation in any sense. Instead, a Kantian theory of manipulation underlines that every person can (and ought to) contribute to her own security from manipulation.

This leads us to my second claim. One must strive for personal autonomy in order to protect oneself from the deteriorating effects of manipulation. While there are other attempts on building a Kantian conception of personal autonomy (see Taylor, 2005; or Formoza, 2013), I do it in a quite different manner. I argue that from a Kantian perspective

personal autonomy can be established through adoption of Kant's three maxims of common understanding. I also argue that these three maxims are in fact three different formulations of one and the same maxim. Adopting it involves a very sophisticated thought experiment which helps us partially to protect ourselves from the deteriorating effects of manipulation.

In the first chapter I will show how empirical psychology fits into the Kantian framework. In the second chapter, I will present some parts of Kant's account of empirical psychology. I will focus on the empirical mechanisms of human action and cognition and defects in cognition and volition. In the third chapter, I will present a brief account of Kant's natural teleology to show how the nature, in a sense, manipulates human beings into developing the culture, making sociability and progress possible. In the fourth chapter, I will show that the very same features of human nature that, according to Kant, make sociability and progress possible are responsible for our susceptibility to manipulation. I will also discuss here particular manipulative techniques and the ways in which they exploit defects in cognition and volition. In the fifth chapter, I will look into how Kant understands relations between autonomy and enlightenment in order to understand whether Kant's idea of autonomy can contribute somehow to our defense against manipulation. I will also touch upon the issue of moral responsibility for manipulation. Finally, in the sixth chapter I will show how exactly we can use the maxims of common understanding to protect ourselves from manipulation.

1 On the possibility of a Kantian empirical psychology

In this chapter I am going to look into whether there is a possibility for empirical psychology as a science within a Kantian framework. I will start by explaining why this inquiry is important for this thesis.

I define manipulation as a deliberate action that aims to change the behavior or beliefs through **conscious** lying or **conscious** usage of techniques that preclude or hinder a moral deliberation. This concept of manipulation presumes that a manipulator uses his techniques consciously. For that reason I regard a manipulator as a person who has set certain ends and uses manipulative techniques in order to achieve them.

The manipulative techniques should be efficient enough; otherwise there is no point in engaging into them. The efficiency of the manipulative techniques implies that there should be some cause-and-effect relations involved. The manipulator should have knowledge of these relations, if he wants to be successful. What kind of relations can it be?

The end of the manipulator is to change beliefs and/or desires of a manipulatee, hence he needs the knowledge about cause-and-effect relations behind the behavior and mind of human beings, which is basically a subject of empirical psychology. It is not necessary for the manipulator to have a highly systemized and refined knowledge about these relations; it suffices if he is just aware that, say, this particular manipulative technique may influence people in that particular way. In other words, he needs knowledge-that, rather than knowledge-how. Still, if empirical psychology as a system of knowledge is not possible within a Kantian framework, then the Kantian framework is not suited for the explaining of how manipulation works and what one can do to defend oneself from it.

The problem is that Kant's attitude towards the empirical psychology is not unequivocal. In *Metaphysical Foundations of Natural Science* he puts it rather bluntly that "[t]he empirical doctrine of the soul can ... never become ... a science of the soul, nor even a psychological experimental doctrine" (Kant, 4:471)¹. This claim seems to be consistent with the importance that Kant puts on freedom as an uncaused cause and absolute spontaneity in his moral philosophy. According to Kant, the moral worth of a good action is derivative; it is derived from the fact that the will behind this action is good (see Kant, 4:393-394). The will is good only when it is fully determined by the moral law rather than by instincts or inclinations. Any trace of natural causation within the human action may compromise its moral worth. If I am doing something which seems morally good not out of respect for the moral law, but just because I have a psychological inclination towards this kind of behavior, then my action lacks moral worth². What makes a person essentially free is that she is the legislator of the moral law. She "is subject only to laws given by herself" (Kant, 4:332). Since for Kant, 'ought' presupposes 'can' (cf. Kant, 3:A548/B576), this picture of moral freedom supposedly implies that a moral agent can exempt herself from the causal effects of her inclinations, desires, instincts when they are incompatible with the demands of the moral law. Moreover, even when they are in compliance with the moral law, they cannot be a decisive causal power behind the morally good action. Henry E. Allison successfully captures the implications of Kant's transcendental freedom in his 'incorporation thesis': "the intentional actions of a rational agent are never 'merely' the causal consequences of ... any ... antecedent conditions", they necessarily require "an act of spontaneity" (Allison, 1990, 5). In other words, desires

¹ Citations from the Critique of Pure Reason are located by reference to the pagination of Kant's first 'A' and/or second 'B' editions. All other passages from Kant's works are cited by the volume and page number, given by Arabic numerals separated by a colon, in the Academy edition of Kant's writings.

² However, it does not follow from this that I cannot simultaneously have such an inclination and still be determined by the respect for the moral law.

(as empirical mental states) can determine a rational agent's will only if she allows them to do so³.

From this perspective it is hard to understand the place of empirical psychology within a Kantian framework. It may seem that Kant's understanding of freedom is akin to Sartre's one, in that "there is no human nature" (Sartre, 1956, 290). Or even if there is human nature, it is not that important to understand or study it, because a rational agent is able to override all its effects. However, it is definitely not Kant's position, because he believes both that there is human nature and that its study is incredibly important. But why is it important? First of all, because human beings are not perfectly rational. There are "the subjective conditions in human nature that hinder people or help them in the carrying out of the laws of a metaphysics of morals" (Kant, 6:217). Hence, it is crucial to understand these conditions for the realization of morality by human beings. Kant clearly acknowledges that ethics should have the rational part and the empirical part. He calls the latter practical anthropology (see Kant, 4:388) or moral anthropology (Kant, 6:217); the main function of it is "to gain for [moral laws] access to the human will" (Kant, 4:389). Moral anthropology, however, should not be confused with what Kant calls anthropology from a pragmatic point of view⁴. The subject of pragmatic anthropology is wider than that of moral anthropology. It is not limited to the purely moral matters, but concerns also the technical and prudential applications of the anthropological knowledge (see Wood, 1999, 203-205).

While pragmatic anthropology and empirical psychology are not the same thing⁵, the former presupposes the possibility of the latter. I am not going to elaborate on this much, but what is important is that empirical psychology is a purely theoretical enterprise, while pragmatic anthropology is, in a sense, an applied discipline regarding a human being as capable of goal-setting. Pragmatic anthropology involves "the investigation of what [a human] as a free-acting being makes of himself, or can and should make of himself" (Kant, 7:119). Kant also states that pragmatic anthropology teaches us how "to use other human beings skillfully for one's purposes" (Kant, 7:322).

There are a few other features of pragmatic anthropology we need to mention. First, Kant regards pragmatic anthropology as a popular discipline. In other words, it is not only for the members of academia, but for a broad audience. Its aim is to promote

³ Some commentators even ascribe to Kant a view that it is impossible to explain the intentional actions in empirical terms. For a further discussion of these positions see Frierson, 2014, 10-14; Loudén, 2000, 17-8; Cohen, 2009, 31-32; Wood, 1999, 178-180.

⁴ In the following I will refer to it as 'pragmatic anthropology'.

⁵ For a discussion of relations between anthropology and psychology in Kant see Wood, 1999, 197-198; Loudén, 2011, 79; Frierson, 2014, 43-49.

“enlightenment for common life” (Kant, 25:853). Second, it is the general knowledge of the world (Weltkenntniß), i.e. it concerns “the nature of humanity, not the state of human beings, for the particular properties of human beings always change, but the nature of humanity does not” (Kant, 25:471). However, in spite of this, pragmatic anthropology has sections focused on the character of the sexes and the different nations.

As Robert Louden stresses, “pragmatic anthropology can be put to many different purposes, some of which are blatantly immoral”, for example, “shrewd politicians may exploit their knowledge of human nature to advance their own personal agendas for power and control” (Louden, 2011, 69). It seems that pragmatic anthropology can explain how a manipulator succeeds in using other humans for his ends. But there is still an open question how is that kind of knowledge is consistent with other parts of Kant’s philosophical system. It is being debated vehemently, and I am not intending to contribute to this debate in this chapter, as it is beyond the scope of this project. In the following I am rather going to briefly sketch one possible way of reconciling Kant’s empirical psychology with the transcendental freedom. There are two perspectives on the human action:

a rational being has ... two standpoints, from which it can consider itself and cognize the laws for the use of its powers, consequently all its actions: first, insofar as it belongs to the world of sense, under natural laws (heteronomy), and second, as belonging to the intelligible world, under laws which are independent of nature, not empirical, but rather grounded merely in reason. (Kant, 4:452).

The second perspective is practical. According to it every rational agent is essentially free; it “is the proper standpoint for moral philosophy” (Frierson, 2014, 14). According to the first perspective, every human action is a part of the empirical world, i.e. it is a phenomenon; and as any other phenomenon it is determined by the causal laws. However, there is a fundamental incompleteness in any causal explanation. “For any causal law of human thought or action, one can always ask why that law has the structure that it does”. While “[t]his possibility does not imply that these causal laws are any less natural, nor any less predictive” (Frierson, 2014, 15), it leaves some place for transcendental freedom, hence empirical psychology is not necessary inconsistent within a Kantian framework. I do not need to argue for a stronger claim here.

There is also another reason why Kant states that a genuine “science of the soul” is not possible. It is related to Kant’s understanding of the term ‘science’. To put it simply, his problem with psychology is that it lacks a priori basis and cannot be the source of apodictic truths. In other words, even if there are noticeable regularities, we cannot describe them in the absolutely universal terms as they are essentially contingent. Still,

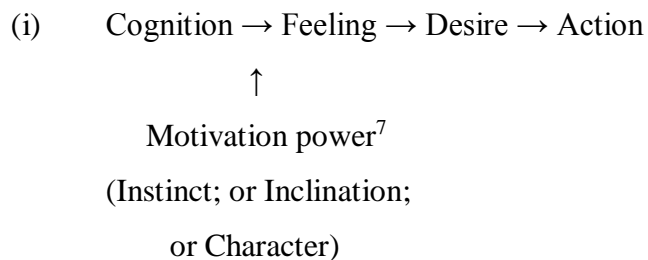
Kant allows that psychology is possible as a “historical systematic natural doctrine of the inner sense” (Kant, 4:471), which is good enough for my goals in this project. A manipulative technique should not have a guaranteed effect in order to be effective.

2 A brief account of Kant’s empirical psychology

In this chapter I will present some parts of Kant’s empirical psychology, namely I will focus on the empirical accounts of human action and cognition and defects of cognition and volition. This will help to explain the mechanism of manipulation⁶. I will here substantially rely on Frierson’s systematic reconstruction of Kant’s empirical psychology (see Frierson, 2005; Frierson, 2014).

2.1 Human action

As I have said in the previous chapter, a manipulator should have some knowledge of cause-and-effect relations behind the behavior of human beings if she wants to be successful in achieving her ends. The possibility of manipulation can be explained only by tracing the series of causes and effects. And it brings us to Kant’s account of human action. Frierson sums this account up thus: “cognition of an object gives rise (sometimes) to a feeling of pleasure or pain, and that feeling gives rise (again, sometimes) to a desire or aversion for the object” (Frierson, 2014, 56). Hence, the most basic model of action looks this way:



Cognition, feeling and desire correspond to the three faculties of the mind (see Kant, 5:198). In addition to this threefold distinction Kant distinguishes between higher and lower faculties. The lower faculties are passive and receptive, while the higher

⁶ I will present it in the fourth chapter.

⁷ Motivational power is the mechanism which carries out the transition from ‘cognition’ to ‘feeling/desire’. Technically, it belongs to faculty of desire; i.e. inclinations and instincts are motivational powers of the lower faculty of desire, while character is a motivational power of the higher faculty of desire.

faculties have relative spontaneity. I use the word ‘relative’ in order to distinguish it from the absolute spontaneity of the transcendental freedom. To make things clearer we can relate it to what Kant calls “a comparative concept of freedom”; according to it “something is called free action if its determining natural basis lies in the acting being internally” (Kant, 5:96). The higher faculties are determined by their internal organization, which gives a human being a sense of psychological freedom⁸ when he is using them.

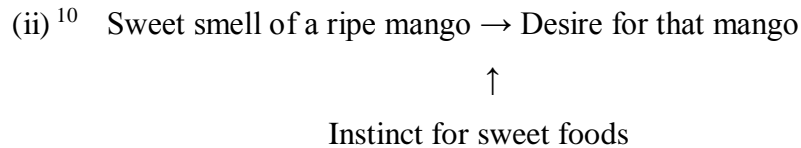
Kant refers to the lower faculty of cognition as sensibility; it includes the senses and imagination. The higher faculty of cognition is called understanding (in a broad sense); it includes three cognitive powers: the understanding in a narrower sense, reason and the power of judgment (see Kant, 7:196). Kant makes a similar distinction between lower and higher faculties of feeling and desire. But their explanation is derivative. If one has a purely sensible cognition, then it may give a rise to a sensible feeling of pleasure (or displeasure) and, then, to a sensitive desire. However, if the initial cognition involves the understanding in a broad sense, then the resulting desire will fall under the higher faculty of desire; i.e., in Kantian terms, it will be intellectual. The content of the desire is irrelevant for this classification; what matters is whether the higher faculty of cognition is involved in giving rise to this desire or not.

I will omit the discussion of the transition to an ‘action’ phase, because having a desire necessarily leads to the agent’s attempt to change the world somehow. If it does not lead to this, then it, according to Kant’s terminology, is merely a wish. For example, if John has two conflicting desires, and one of these desires is stronger and, as result, John is moved by it, then we should classify his unrealized desire not as a real desire, but as a wish. In other words, a desire – as Kant defines it – always has practical consequences; it always leads to some action.

I am also not going to pay too much attention to the ‘feeling’ phase and the transition from it to the ‘desire’ phase. There are two kinds of feelings of pleasure (or displeasure): aesthetic and practical. The pure aesthetic pleasure never gives rise to a desire; whereas the practical pleasure “is the direct cause of desire” (Frierson, 2014, 59). The connection between the practical feeling and desire usually does not need any additional investigation. Affects are the only exception, but I will consider them later. What is of interest to me here now is how the transition from cognition to pleasure/desire works:

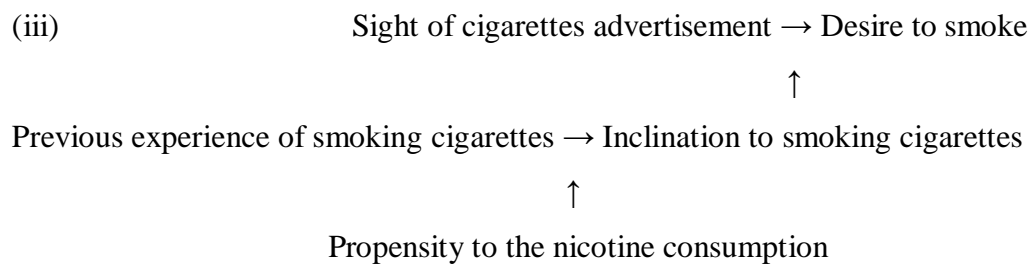
⁸ It means that an agent feels that he was not coerced into doing what he has done, that he could have done otherwise.

a) In case of the lower faculty of cognition, this transition is explained by appeal to either instincts or inclinations. Instincts are natural predispositions. They do not have a further causal explanation⁹; they are innate, and living beings (including humans) just have them. Here is the example of how an instinct explains transition from the cognition to the desire:



In this example a sensory cognition of a mango gives the rise to the desire for the mango, and this connection between cognition and desire is explained by the innate instinct for sweet foods.

Unlike instincts, inclinations are not natural, but acquired (see Frierson, 2014, 69). They presume some natural predispositions. Any acquired addiction gives us a good example of how the inclination develops.



The inclination to smoking is not innate. In order to develop it, a person must have the past experience of smoking. But in order to explain why this experience may lead to developing an inclination we need additional causal explanation – propensity. “Propensity ... is the inner possibility of an inclination, i.e. the natural predisposition to the inclination” (Kant, 25:1111-1112). There is no need for further explanation, because propensities are natural and innate, just as instincts.

b) In case of the higher faculty of cognition, cognitions that may give rise to desire are practical principles or maxims¹¹. The transition from a maxim to pleasure/desire is explained by appeal to character. While many living beings have instincts and inclinations,

⁹ But they can be explained teleologically. However, it is not relevant to the goals of this chapter; hence I am not going to elaborate on teleological explanation here.

¹⁰ The diagram is taken from Frierson, 2014, 68.

¹¹ A maxim here should be understood as any principle that governs action. It should not be necessary a moral principle. A moral principle is a specific maxim which can be universalized.

only rational beings have character. It should be noted that Kant uses the term ‘character’ for the different purposes. Sometimes he uses it to refer to the intelligible character of a person as noumenon. In that case the character “is not itself appearance” (Kant, 3:A539/B567) and, hence cannot be appealed to in empirical explanation. In this chapter, however, I am talking about what Kant calls ‘character simply’ and defines as “that property of the will by which the subject binds himself to definite practical principles” (Kant, 7:292)¹². In other words, whether cognizing this particular maxim gives a rise to a desire or not depends on one’s character. ‘Cognizing’ in this case does not necessarily mean ‘judging that’. A person can entertain a certain principle without attributing any value to it. This entertaining should not necessary be very explicit. When someone is in the certain circumstances, he considers very quickly a lot of different maxims which can be applied to this particular situation, i.e. he cognizes these maxims.

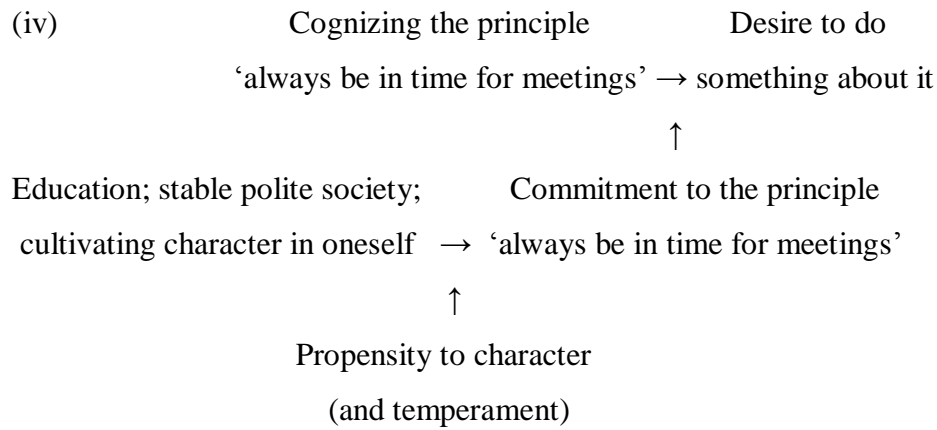
For example: David has overslept and got up later than he had planned, and now he risks missing a meeting. If David thinks about the principle ‘always be in time for meetings’ and he has a commitment to it in his character, then it must have some influence on his actions. Perhaps, he would decide to miss his breakfast in order to immediately get a taxi. But if David does not have a commitment to this principle, he may think something like: “Well, some people may believe that it is important to always be in time for meetings; but I think that being late is not that bad”. In that case he entertains the relevant maxim, but there is no a commitment to it in his character, hence it does not give rise to a desire.

Character is akin to inclinations in that it is not a natural disposition, but acquired (see Kant 7:294). As with inclinations, there is a propensity to character which is innate and natural. Kant also states that other factors may contribute to developing of character. For example, different temperaments¹³ have an influence on it (see Kant, 25:1388). But what is more relevant for this thesis is that “[t]he acquisition of good character with people happens through education” (Kant, 25:1172) and its development is influenced by “various social institutions ... including stable and just political regimes, peace, and even progress in the arts and sciences” (Frierson, 2014, 79). According to Kant, a person herself “through understanding and reason” (Kant, 25:1172) can also cultivate her character.

It can be summed up thus:

¹² For a brief account of different Kant’s uses of the term ‘character’ see Frierson, 2014, 73.

¹³ Temperament for Kant is also a natural predisposition and does not need any further causal explanation.



As we can see, according to Kant, the practical principles are crucial part of rational action. The investigation into the mechanism of their formation will be my goal in the next sub-section.

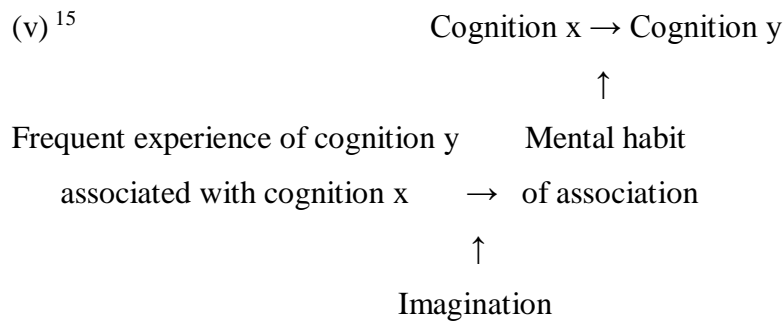
2.2 Human cognition

The human beings have higher and lower variations of the faculty of cognition. The lower faculty of cognition is sensibility; it includes the senses and imagination. The higher faculty of cognition is called understanding (in a broad sense); it includes three cognitive powers: the understanding in a narrower sense, reason and the power of judgment (see Kant, 7:196). I am not going to present here Kant’s empirical account of the senses, because it is not that relevant for my thesis¹⁴. Imagination, on the other hand, plays a very important role in producing prejudices and causing enthusiasm. Both prejudices and enthusiasm (as an affect) are used by a manipulator in order to bypass a manipulatee’s rational deliberation.

a) Imagination is a “faculty of intuition without the presence of the object” (Kant, 7:167) It can be productive or reproductive. If I remember something that I have experienced before, I use reproductive imagination (which is, basically, memory). But when I construct something new in my imagination, then I use its productive form. However, even productive imagination “is not capable of producing a sense representation that was never given to our faculty of sense; one can always furnish evidence of the material of its ideas” (Kant, 7:168). In other words, it combines cognitions or simple components of cognitions we have already experienced into new ones. It can create connections between our cognitions. The way Kant treats imagination strongly resembles Hume’s account of imagination (see Hume, 1978, 283). Kant posits the law of association,

¹⁴ For Kant’s empirical account of the senses see Frierson, 2014, 94-96.

according to it “empirical ideas that have frequently followed one another produce a habit in the mind such that when one idea is produced, the other also comes into being” (Kant, 7:176). “Association is based on three elements, on accompaniment, contiguity, and on relation” (Kant, 25: 512). Here is the basic model:



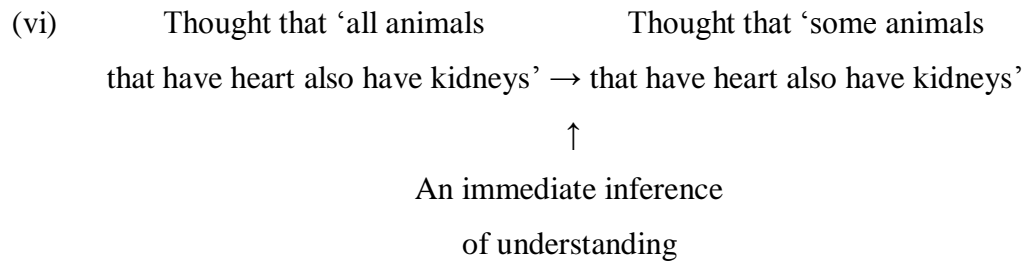
This function of imagination is incredibly important; it explains, among other things, the connection between any word and corresponding concept. However, as I will show later, imagination may also lead to the development of unsound provisional principles, or prejudices.

b) The higher faculty of cognition is divided into three cognitive powers: the understanding, reason and the power of judgment. Each of these powers has its own principles governing transition from one cognition (or a set of cognitions) to another. The understanding carries out a transition from a set of sensory cognitions to the concept. This process involves three stages: comparison, reflection, and abstraction (see Kant, 9:95). E.g. I see a lot of different animals, compare them, reflect on what is common between them, and abstract from it everything else, as result, thus I form a concept of an animal in general.

All three powers govern transitions between judgments, and each power has its own principles. These principles are called provisional; they are “maxims for the investigation of a thing” (Kant, 9:75). The principles of reason are basic syllogistic forms, such as modus ponens and modus tollens (see Kant, 9:130). The principles of the power of judgment are analogy (an inference from particular to total similarity of two things) and induction (an inference from the particular to the universal in regard to many objects) (see Kant, 9:133)¹⁶. And the principle of understanding is an ‘immediate inference’ from the universal to the particular. For example:

¹⁵ The diagram is taken from Frieson, 2014, 98.

¹⁶ It should be noted, that the principles of the power of judgment cannot produce infallible knowledge.



It was an account of a how the higher faculty of cognition ought to work. But, in practice, it is often influenced by sensibility and, thus tends to carry out the transitions between judgments on the basis of unsound provisional principles, or prejudices. In the next sub-section I will consider the defects of cognition and volition and briefly discuss their sources.

2.3 Defects of cognition and volition

In this sub-section I am not going to discuss all possible defects of cognition and volition. I will focus on those which can be exploited by a manipulator. More specifically, I will consider prejudices, affects and passion.

a) In addition to prejudices there are other cognitive defects, such as mental and cognitive disorders. But I believe it is safe to say that such defects are exceptional cases¹⁷, while prejudices are common and widespread. Kant defines a prejudice as “a principium for judging, not from objective grounds, but ... [from] subjective ones” (Kant, 24:864). First of all, it must be stressed that Kant does not regard the prejudice as a merely false judgment. For him the prejudice is a provisional principle akin to those of understanding, reason and the power of judgment. However, unlike them, the prejudice is an unsound principle; of course, it may lead to the right judgment, but it can never justify it (see Kant, 9:75-76). I will give few examples of prejudices to make things clearer:

- A logical egoism is an unsound principle, according to which your judgment is better than judgments of others just on the ground of being yours.
- The prejudice of the prestige of the multitude (the multitude prejudice) is an unsound principle, according to which the judgment shared by majority is better than the judgment shared by minority.
- The prejudice of the prestige of the person (the person prejudice) is an unsound principle, according to which the judgment of a certain person is

¹⁷ Surely a manipulator can exploit these exceptional cognitive defects too for his purposes. However, in this thesis I am more interested in why relatively mentally healthy people are susceptible to manipulation.

better than judgments of others just on the ground of being a judgment of that particular person.

These prejudices are highly efficient for manipulation. If a manipulator succeeds in convincing a manipulatee that some judgment is shared by majority or comes from a well-respected expert, and if the manipulatee adopts the prejudices not only as a provisional judgments, but also as provisional principles¹⁸, then it would lead to the manipulatee's acceptance of the judgment imposed by the manipulator.

I want to say few additional things about the person prejudice. It seems perfectly rational to value the judgments of a certain person higher than the judgments of others, if this person is considered to be more competent, i.e. if she is a well-known expert. There are different ways how we can understand why Kant calls it a prejudice. First, it can be understood that we should not trust experts in all things, especially in those, which are outside of their area of expertise. For example, the fact that Chris is a well-known expert in quantum physics does not make him a competent specialist in linguistics. Hence, it would be epistemically wrong to appeal to Chris's reputation of being a great scientist, when the matter of discussion is beyond his competence. Second, it can be understood that there are certain fields of knowledge where one's prestige cannot justify anything. Morality is a good example. All moral truths are a priori, and they are equally accessible for all rational beings. Hence, from a Kantian perspective, there is no such thing as a genuine expert in morality. Third, we may understand it conditionally. The fact that some claim was made by an expert should not be taken as a sufficient justification for validity of this claim, because (1) the expert can be fake¹⁹; and (2) even if he is a real expert, he may lie to us. If we are absolutely sure that (1) and (2) are false, than it would be rational to believe the expert. The problem is that both (1) and (2) are very difficult to rule out. Furthermore, a critical attitude towards experts is also a matter of self-respect. In 'Answering the Question: What is Enlightenment?' Kant compares a thoughtless trust in experts (guardians) with immaturity:

It is so comfortable to be immature. If I have a book that reasons for me, a pastor who acts as my conscience, a physician who determines my diet for me, etc., then I need not make any effort myself. It is not necessary that I think if I can just pay; others will take such irksome business upon themselves for me. The guardians who have kindly assumed supervisory responsibility have ensured

¹⁸ The difference is that a prejudice may lead to the right judgment. Hence, provisional judgments may have some provisional or heuristic merit. However, it would be wrong to think that the rightness of the judgment follows from the prejudice. In other words, the prejudice cannot justify the right judgment. One may use provisional judgments without adopting them as principles. In that case, he is less vulnerable to manipulation, because he does not regard new cognitions as a real knowledge (i.e. based on objective grounds), but rather as an opinion or hypothesis (i.e. based on subjective grounds).

¹⁹ I will elaborate more on how the multitude prejudice may lead to the person prejudice in the fourth chapter.

b) Affects and passions are similar to prejudices in that they also bypass rational deliberation. Kant refers to them as “illness[es] of the mind, because both affect and passion shut out the sovereignty of reason” (Kant, 7:252). Kant defines an affect as the disorder of the faculty of feeling; but he also uses the word ‘affect’ to refer to a particular feeling which leads to this disorder. As a feeling it does not involve any perspective in the future, it is here and now. A person does not deliberate on what she is going to do or how she is going to do it; an affect almost immediately leads to some action or inactivity.

For example: Josef says something incredibly offensive to Ceferino. Ceferino immediately punches Josef in the face. Ceferino does not have a commitment to a practical principle ‘always punch those who say something offensive to you’. Moreover, Ceferino is a pacifist; he has a commitment to the principle ‘never hurt physically anyone’. But an affect made him unable to behave in accordance with his principles. Perhaps, few minutes later Ceferino even would ask Josef for forgiveness for being excessively violent.

Kant stresses that “it is not the intensity of a certain feeling that constitutes the affected state, but the lack of reflection in comparing this feeling with the sum of all feelings (of pleasure or displeasure)” (Kant, 7:254). Other feelings and corresponding desires just end up not being taken into account. Ceferino was so overwhelmed by the painful feeling of anger that he was not even able to think about his principles.

c) Passions are similar to affects in that they preclude rational deliberation. However, passions allow at least some reflection. For Kant, a passion is the disorder of the faculty of desire, but he also uses the word ‘passion’ to refer to a particular desire which is constitutive of this disorder. Frierson argues that we should understand ‘passion’ as a “practical principle that has sensuously given ends” (Frierson, 2014, 228-230) and that precludes any comparing of it with the sum of all desires in respect to a certain choice.

For example: Mani has said something incredibly offensive to Luke. Luke has not showed any reaction to this, but he has decided to revenge Mani in the future. Basically, he has adopted a principle ‘revenge is a dish best served cold’. In the future, Luke can be very inventive in following this maxim; he can deliberate a lot on the sub-ordinate maxims in order to invent the most ingenious way of revenge. “Passion always presupposes a maxim on the part of the subject, to act according to an end prescribed to him by his inclination” (Kant, 7:266). Hence, his rational deliberation is not fully precluded. However, if it is a real passion, then Luke would find it extremely hard to drop it, as he is unable to compare it with other desires. Basically, his passion precludes reflection on happiness or morals. A passionate person also tends to rationalize his passion. For instance, Luke may convince himself that his mania for vengeance, actually, is not a passion, but a moral duty.

According to Kant, this rationalization is primarily intended to repress one's sense of guilt; Kant calls it self-deception. Self-deception is a particular example of the secret influence of sensibility on the higher faculty of cognition.

Affects and passions interfere with normal processes of rational choice. They can be exploited by a manipulator. The manipulator can – by provoking an affective or passionate state²¹ in a manipulatee – trick him into doing something that is somehow advantageous for the manipulator. Affects corrupt the faculty of feeling, compromising possibility of choice; and passions corrupt the faculty of desire or choice itself.

I will say few words on how one can cope with affects and passions. The susceptibility to affects shows a lack of virtue²². If one is already in the affective state, then there is little she can do. However, it is possible to prepare oneself for the future encounters with affects by cultivating virtue of self-mastery that Kant also calls courage. The same can be said about passions; one should cultivate self-mastery to be prepared. But there is also another thing which makes passions extremely difficult to deal with. Unlike affects, many passions are the product of the social development. People acquire them when they begin to compare themselves with others. It gives a rise to such passions as envy, ingratitude, spite (Kant, 6:27), manias for honor, power, and wealth (Kant, 7:268). Passions, at least partially, are a social problem, and they can be pacified by the further socio-cultural development.

What unites prejudices, affects and passions is that all of them involve a secret influence of sensibility on the higher faculty of cognition and, as result, an inability to think for oneself. As I will argue in the other chapters, 'think for oneself', for Kant, is not merely an epistemological principle, but also a demand of the moral law, and all defects of cognition and volition are, in fact, consequences of what Kant calls radical evil of human nature.

3. Nature as an original manipulator

In the previous chapter I have presented some parts of Kant's account empirical psychology which are necessary for the causal explanation of the mechanism of manipulation. In this chapter, I am going to examine human susceptibility to manipulation from another standpoint. I will start this chapter by presenting a brief account of Kant's

²¹ It can be done, for instance, by exposing a manipulatee to images of gruesome murder scenes or war crimes.

²² Virtues are inclinations which help people in carrying out the demands of the moral law. They must be consciously cultivated. According to Kant, it is a wide duty of every human being.

natural teleology. Then I will show that from a teleological point of view (i.e. if we assume that nature has its own ends²³) nature itself can be regarded as a manipulator, because it tricks human beings into developing culture.

This teleological account is helpful for two reasons. First, it shows that the very same features of human nature that, according to Kant, make socio-cultural progress possible are also responsible for human susceptibility to manipulation. In a sense, a manipulator just imitates nature. Second, this account shows that the manipulative techniques are essentially woven into the very fabric of society. It explains why they are not just mechanically possible²⁴, but also wide-spread and efficient.

3.1 Kant's natural teleology

In the *Critique of the Power of Judgment* Kant says that some objects or actions can be understood and explained only if it is assumed that there is a will with the certain ends and principles behind those objects or actions. In Kantian terms, these objects or actions are called purposive.

An object or a state of mind or even an action, however, even if its possibility does not necessarily presuppose the representation of an end, is called purposive merely because its possibility can only be explained and conceived by us insofar as we assume as its ground a causality in accordance with ends, i.e., a will that has arranged it so in accordance with the representation of a certain rule (Kant, 5:220, 105).

To put it simply, something is purposive when it appears as if designed or produced according to the idea of some purpose. Kant distinguishes between two kinds of purposiveness:

1) A subjectively purposive action or object appears “as if designed for the sake of our cognitive capacities”²⁵.

²³ It should be stressed that Kant does not make an ontological claim that nature has its own ends. He states it very clearly that “we do not actually observe ends in nature as intentional, but merely add this concept as a guideline for the power of judgment in reflection on the products of nature” (Kant, 5:399).

²⁴ The mechanical possibility of manipulation is explained in the second chapter.

²⁵ Subjective purposiveness can be of two subkinds - logical and aesthetic. Logical purposiveness is displayed by the activity of the cognitive faculties in the perception of nature as a whole. According to Kant, we tend to – and, actually, ought to – regard nature “as doing us an epistemological favor by making possible both its taxonomical ordering in terms of a coherent set of concepts and its nomological ordering in terms of a set of empirical laws that allow for the construction of overarching theories” (Allison, 2009, 30). Aesthetic purposiveness is displayed by the activity of the cognitive faculties in the perception of the beautiful. The pure judgments of beauty involve a free play of imagination and understanding, this free play cause us regard beautiful objects as formally purposive. While it is a very interesting topic which is not fully irrelevant to the problem of manipulation, I will not elaborate further on aesthetic purposiveness here, because it is beyond the scope of this thesis.

2) An objectively purposive action or object appears “as if designed with respect to [its] own inner possibility” (Alisson, 2009, 29). In the following I will focus exclusively on objective purposiveness.

Objective purposiveness is the principle of teleological judgment, i.e. when one judges something teleologically, he hypothetically presupposes that it serves some purpose. This purpose can be either internal or relative. For example, according to Kant, every organic being should be considered as a natural end, i.e. as internally purposive, because functions of its parts can be explained only in relation to the whole. Hence, the idea of the whole organism conceptually predates its own parts (see Kant, 5:376). While the notion of a natural end is a very important for Kant’s philosophy of biology, I am more interested here in another form of objective purposiveness – a relative one. Kant makes a rather strong claim that assuming internal purposiveness of organisms “necessarily leads to the idea of the whole of nature as a system in accordance with the rule of ends” (Kant, 5:379)²⁶. In other words, we should (or, at least, may) regard nature teleologically, i.e. as a complex hierarchical system where one object is relatively purposive to another. And the word ‘nature’ here should be understood in a very broad way: everything determined by causal mechanical laws is a part of nature.

Kant argues that if we consider nature this way, we shall see that everything exists for the sake of humankind (see Kant, 5:427). What makes a human being unique is that “he is the only being on earth who forms a concept of ends for himself and who by means of his reason can make a system of ends out of an aggregate of purposively formed things” (Kant, 5:427). Kant also assumes that there should be a final end of creation, which is something absolutely unconditioned, something that is good for its own sake (see Kant, 5:426). There are two candidates for the final end: morality and happiness. Wood interprets Kant’s notion of the final end “as having two components related ... conditionally” (Wood, 1999, 312), i.e. morality and happiness proportional to morality. Morality from a Kantian perspective presupposes transcendental freedom, i.e. such a will that is not fully determined by causal mechanical laws. In other words, morality cannot be a product of nature. Hence, the final end has to be outside of nature, it can be set only by a free rational being as noumenon. The role of nature, i.e. its ultimate end²⁷ is to prepare a rational being for exercising freedom of choice. It can be done through bringing culture. Kant distinguishes between two forms of culture: the culture of discipline which Kant defines as

²⁶ However, literally on the next page Kant makes a much weaker claim that “we **may** go further” (Kant, 5:380) (bold emphasis mine). I tend to agree with Allison who thinks that the reference to necessity should be ignored (see Allison, 2009, 36).

²⁷ The ultimate end of nature and the final end of creation should not be confused!

negative and consisting in “the liberation of the will from the despotism of desires” and the culture of skill which is understood as “[t]he production of the aptitude of a rational being for any ends in general (thus those of his freedom)” (Kant, 5:431-432). It is the second form of culture which is especially crucial for transition from nature to freedom.

To sum it up: the ultimate end of nature considered as a teleological system is to make humankind cultured (in a very specific sense). By doing this nature prepares humans for exercising freedom of choice. Then humans may set up a final end which is morality and strive towards it. I want to stress two things here. First, Kant does not claim that nature **is** a system of purposes which has an ultimate end. Objective purposiveness is only “a principle for guiding the investigation” (Kant, 20:236); it does not lead to any ontological commitments. Second, the connection between the ultimate end (culture) and the final end (morality) belongs not to nature, but freedom. It means that this connection is not necessary. It is perfectly possible, that a cultured person may never set morality as her end. Kant is not a historical determinist like Hegel; his position is closer to that of Rousseau²⁸. He does not treat moral progress as something inevitable. In *The Conflict of the Faculties* Kant directly says that “we are not capable of placing ourselves in this position when it is a question of the prediction of free actions ... these actions, of course, the human being can see, but not foresee with certitude” (Kant, 7: 84).

3.2 The unsocial sociability

As I mentioned in the second chapter, for Kant, all defects of cognition and volition are consequences of radical evil of human nature. And from teleological point of view it seems that all qualities of human nature should be apt to contribute somehow to socio-cultural progress²⁹. But how can the propensity to evil contribute to socio-cultural progress?

To begin with, I briefly explain how Kant understands the evil. Human beings have three basic predispositions to good: animality, personality, and humanity³⁰. The predisposition to animality is what Kant calls a mechanical self-love. It manifests itself in the drive for self-preservation, propagation, and communication with others. The

²⁸ As Frederick Neuhouser argues “Rousseau’s account of evil shows that, if we can have no guarantee of there being a way out of our present fallenness, we can also not know a priori that no such path exists” (Neuhouser, 2008, 8). However, we can see that Rousseau’s agnosticism with respect to the possibility of moral progress is noticeably more pessimistic.

²⁹ These qualities are determined by causal mechanical laws; hence, they are the part of nature.

³⁰ According to Kant, they are natural, not acquired. Hence, they should be treated akin to instincts. But it should be noted, that other animals have only the first predisposition and in the limited form. Predispositions to humanity and personality are uniquely human.

predisposition to personality is “the susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice” (Kant, 6:27). It explains why the demands of the moral law may influence one’s behavior. The predisposition to humanity consists in the comparative self-love. This predisposition gives rise to “the inclination to gain worth in the opinion of others” (Kant, 6:27). Basically, it is a drive for recognition. We can find a very similar notion of amour-propre (French, "self-love") in Rousseau’s thought. Rousseau describes it as an exclusively human sentiment that involves comparisons with others and seeing oneself as others see one.

All three predispositions, considered in themselves, are good. The evil arises only when the incentives of two kinds of self-love are prioritized to those of moral law:

Whether the human being is good or evil, must not lie in the difference between the incentives [of self-love and morality] that he incorporates into his maxim ... but in their subordination (in the form of the maxim): which of the two he makes the condition of the other. It follows that the human being (even the best) is evil only because he reverses the moral order of his incentives in incorporating them into his maxims (Kant, 6:36).

Simply stated, the evil consists in a failure to make the moral law the ultimate condition of all other maxims. As we can see, Kant’s understanding of the evil is very broad. Even the noblest deeds would be manifestations of our propensity to evil, if they are done for the wrong reason. It should be noted, that Kant defines it only as a propensity, i.e. as “the subjective ground of the possibility of an inclination” (Kant, 6:29). It means that it is not logically impossible for a human to never develop an inclination to evil. It is just highly improbable, if one is in the social condition, where the incentives of the comparative self-love are very strong (see Wood, 1999, 288). However, exactly these incentives are crucial for explanation how the propensity to evil can contribute to socio-cultural progress and, hence to one’s ability to exercise freedom of choice.

Since comparative self-love makes one attach more importance to oneself than to others and to demand the same consideration from others, it is the source of never-ending antagonism between people. In the essay ‘Idea for a Universal History’ Kant shows a strong connection between antagonism and the development of culture. He argues that without antagonism human beings would not be very different from the sheep they herd. Antagonism causes what Kant oxymoronically calls the unsocial sociability. The unsocial sociability has two conflicting components:

1. The inclination to be associated with others. A human alone is not able to satisfy all his needs; hence he has to cooperate with other people. He also feels himself “to be more human” (Kant, 8:21) in society, he can achieve

more there and it is easier to defend these new achievements with the help of others.

2. The inclination to be isolated from others. A human is always on the alert. He is inclined to arrange everything according to his own goals, thus abusing the freedom of other people, hence he is expecting the same thing from them. This attitude helps him to mobilize all his powers and suppress his laziness³¹ in order to “establish a position for themselves among [his] fellows” (Kant, 8:21).

The drive for recognition stimulates competitiveness between people. As Schneewind nicely puts it, “[t]he energies we devote to showing others how much stronger and smarter we are lead us to create ingenious inventions and brilliant new ideas that gradually enrich and enlighten our strife-ridden common lives” (Schneewind, 2010, 320). But in order to get along people – who perceive each other as competitors – often have to conceal their real thoughts and feelings. This necessity stimulates the sense of propriety which is “an inclination to inspire the respect of others toward our persons through good manners (the hiding of that which could arouse disdain)” (Kant, 8:113)³². The sense of propriety is double-edged: on the one hand, a person by exercising propriety can trick others into respecting her; on the other hand, this person herself may be considered as a victim here, because in order to gain an approval of others she has to conform to received customs. As Kant says, “[o]n the whole, the more civilized human beings are, the more they are actors. They adopt the illusion of affection, of respect for others, of modesty, and of unselfishness” (Kant, 7:151).

There is a similar antagonism on the international level. All states want to practice unrestricted freedom; therefore, they are expecting the abuse of freedom from the other states. Never-ending wars and war preparations are the great burden for every state. At the same time, the danger of war helps to sustain the level of culture and inner civil liberties, because any encroachment on them may weaken economy and, hence the military power of the state. In the future, the costs of the antagonism among the states would get bigger,

³¹ Laziness is primarily caused by the predisposition to animality, more specifically, by the uncontrolled drive for self-preservation. And the predisposition to humanity, i.e. the drive for recognition helps to overcome it. But there are ways in which the drive for recognition can, instead, contribute to laziness. E.g. sometimes laziness is stimulated by one’s desire to be efficient. Adopting prejudices as the provisional principles that govern transitions between judgments may make one’s productivity higher, but it is still a manifestation of laziness.

³² According to Kant, the origin of the sense of propriety can be found in the control over the sexual instinct. Human beings quickly discovered that the stimulus to sex can be prolonged by the means of imagination. Concealing object of desire (e.g. by fig leaf) makes this stimulus more moderate and, at the time, more enduring and uniform. Thus, the refusal turned merely animal desire to love and feeling “of the merely agreeable over to the taste for beauty” (Kant, 8:113). It is the point where human nature makes the certain kind of manipulation a necessary component of sociability.

and the interconnection among different economies – stronger, thus the defeat of one state would have a deteriorating effect on all others. Eventually, all these factors may³³ force the states to “discover a law of equilibrium ... and to introduce a united power which lends force to this law” (Kant, 8:26). Kant refers here to the establishment of cosmopolitan civil society, which is the precondition for the full development of all human predispositions (see Kant, 8:25), including that to personality. Hence, the antagonism may be regarded as a crucial tool which nature uses to prepare people for freedom.

I argue that, from the teleological point of view, nature, in a sense, tricks people into developing culture. People have their own ends; they are driven by comparative self-love and try to achieve the first and the best rank among others. But by doing this they – perhaps, unwillingly – develop their capacities, and bring about socio-cultural progress. Of course, it should be stressed, that in a strict sense nature is not a manipulator. Because I define manipulation as a deliberate action that aims to change the behavior or beliefs through **conscious** lying or **conscious** usage of techniques that preclude or hinder a moral deliberation; and we do not have sufficient reason to attribute consciousness to nature.

The unsocial sociability could not last without “the secret falsity in even the closest friendship” (Kant, 6:33). Human relations are filled with illusions and pretense, as in order to get along people have to conceal their real thoughts and feelings.

Hence, the manipulative techniques are not just mechanically possible; they are essentially interwoven into the very framework of society. We constantly use some of their elements in our everyday life, for example, when we behave politely to each other. However, there is a very big difference between being a manipulator and being well-mannered. Not all kinds of actions involving illusions or pretense should be regarded as manipulation. I will elaborate on this in the next chapter.

4 The mechanism of manipulation

In the previous chapter, I have shown the role of radical evil of human nature in tricking people into developing culture. In this chapter I will tell how a manipulator may exploit radical evil for his own ends. I will start by going through how the different psychological states make people susceptible to particular manipulative techniques. Then I will discuss how the manipulator can direct public opinion and why people tend to believe in conspiracy theories. At the end of the chapter I will address the difference between

³³ This ‘may’ is important here. According to Kant, it is not necessary that the states will be forced to discover a law of equilibrium. It is just something we ought to hope for and strive for.

manipulation and activities which have similar mechanism, but do not necessary involve misrepresentation or bypassing rational deliberation.

4.1 Conformity

In this section I will show how conformity makes a person susceptible to adopting false beliefs and prejudices. I define conformity as the compliance with prevailing social standards; it is not an inclination, but a certain psychological state, which, however, is primarily caused by an inclination, namely by the sense of propriety.

Kant defines the sense of propriety as “an inclination to inspire the respect of others toward our persons through good manners” (Kant, 8:113). This inclination works as glue for the unsocial sociability. People have to conceal some of their thoughts and feelings if they want to deal with each other. Kant’s account of propriety has connections to treatments of it by other early modern philosophers, particularly by Hume and Rousseau. For Hume propriety as something immediately agreeable to others is a virtuous trait (see Hume, 1978, 591); for Rousseau, it is a social veil on vice and source of evil (see Rousseau, 1973, 6). Unlike Hume, Kant does not think that propriety itself has moral worth. But his position is also quite distinct from the one of Rousseau, as Kant believes that the sense of propriety “gave the first hint toward the formative education of the human being as a moral creature” (Kant, 8:112)³⁴.

At the same time, Kant acknowledges Rousseau’s point that the sense of propriety may make people enslaved by public opinion. In the ‘Letter to d’Alembert’ Rousseau discusses how it can be used for manipulation. In order to make people consider duels to be indecent he proposes to establish a special institution – the Court of Honor – which should have a right of setting rules for warranting and conducting an official duel. Then with time, these rules should gradually become more and more restrictive to make a settlement of conflicts regarding honor free of violence. Rousseau believes that at some moment the proponents of duels – forced by the public opinion – would have to renounce their beliefs as indecent (see Rousseau, 2004).

Kant lists three sources of prejudices: imitation, customs and inclinations. It is easy to see that they are closely connected. People have an inclination to laziness and the sense of propriety (also an inclination); hence, they want to be esteemed by others without doing too much. Imitation of others and observance of received customs is the most

³⁴ I will elaborate on this in the sixth chapter.

straightforward way of achieving this³⁵. Besides, long-lasting imitation may lead to the creation of new custom.

Impact of imitation, customs and inclinations on one's thought and behavior may result in conformity, i.e. the compliance with prevailing social standards. As it involves comparisons with others and seeing oneself as others see one, it may seem that conformity has a similarity with the maxim of a broad-minded way of thinking – to think in the position of everyone else. But this similarity is superficial and there is a fundamental difference. A broad-minded way of thinking is an epistemological principle which helps one to be impartial³⁶, whereas conformity is a passive effect of comparative self-love: one is conformable, because he wants to avoid possible troubles; because he wants to be esteemed by others.

What makes conformity dangerous is that it may lead to an inability to think for oneself; it is a state of potentially compromised autonomy³⁷. Conformity makes the person very susceptible to adopting false beliefs and prejudices. The prejudice of the prestige of the multitude naturally follows from one's conformity. And this prejudice may easily lead to others. For instance: if Ethan adopts the multitude prejudice as his provisional principle, then from the cognition that 'all believe that Mr. Smith is the greatest expert in foreign policy' he may carry out the transition to his own belief that 'Mr. Smith really is the greatest expert in foreign policy', and then from this belief another prejudice naturally follows – that of the prestige of a person.

A manipulator can easily exploit these prejudices; weak-willed conformists are easy targets for him. It should be also noted that the manipulator may try to appeal to non-existing social standards. For example, Sylvia may tell Kerim that everyone believes in P, trying to trick him into believing in P. But let us suppose that, actually, it is not true that everyone believes in P, and Sylvia knows it. In that case Sylvia does not really appeal to public opinion; she just pretends that she appeals to it. Her act of manipulation involves both lying and exploit of the multitude prejudice³⁸.

Still, the influence of real public opinion is very strong, and manipulators know it. It is very hard (at least, psychologically) to sustain autonomy when you are surrounded by people – especially if some of them are close or important to you – who strongly believe in what you consider a total nonsense. Moreover, your belief that you have such autonomy is

³⁵ However, it may be argued that sometimes following customs can be very time- and labor-intensive. In these cases the role of an inclination to laziness is more ambiguous.

³⁶ For more details on the maxim of a broad-minded way of thinking, see the sixth chapter.

³⁷ I will discuss autonomy at length in the next chapter.

³⁸ It also can be argued that the success of her lies depends on another prejudice, say, on the prejudice of unfounded trust. But I am not going to investigate this interpretation further in this thesis.

constantly challenged by the frightening idea that, perhaps, it is you who believes in nonsense, while others are perfectly sane. In any case, your conduct is heavily influenced by the opinion of others. If you are vociferous in your disagreement, those around you will find your conduct outrageous and will try to either help, or force, you to change your beliefs. If you remain silent, then you, de facto, refuse to live according to your own beliefs.

4.2 Ambition and other passions

Whereas conformity is a passive and calming effect of comparative self-love, ambition can be regarded as its active and violent counterpart. In the previous chapter I have discussed how competitiveness stimulates scientific and cultural progress. People mobilize all their powers and suppress their laziness in order to “establish a position for themselves among their fellows” (Kant, 8:21). People want to show that they are better than others; they are moved by the love of honour. Kant even says that without such a love “there would be no incentive to pursue the sciences” (Schmidt, 2016, 43).

But the love for honour may take an unhealthy form. When one is addicted to honour, he becomes ambitious. Ambition can be defined as passion for honour (see Kant, 25:1141-1142). It involves “three ... things: (1) a tendency to think yourself better than others, (2) a desire that others think of you as better than they are, and (3) a desire to be better than they are” (Wood, 1999, 263). This mixture makes people unsocial, hence resistant to the sedative effect of the proprieties. It may even seem that ambition can make people less vulnerable to manipulation, because it often forces them to break with conformity. Moreover, because ambitious people have a tendency to think that they are better than others, they are apt to adopt a very specific prejudice called logical egoism, “which takes as completely dispensable the criterion of truth, to compare one's opinions with those of other men” (Kant, 24:740). While logical egoism is unsound provisional principle, it seems that it may be of help with resisting against manipulation.

However, ambitious people are opened to another kind of manipulation. They are ready to make great sacrifices in order to become esteemed. But the understanding of what makes one esteemed can vary greatly and is still closely tied to public opinion. A manipulator may try to influence this understanding. Once again, it can be done by exploiting prejudices. For example: Daniel is a very ambitious guy, ambitious in a Kantian sense. Nadir says that Daniel should join the military, because it would help Daniel to become esteemed. Nadir can use different ploys to persuade Daniel that joining the

military is the matter of honour and the best way to become esteemed. The most obvious way is to appeal to public opinion.

Actually, I think that if Nadir does not lie, but appeals to real public opinion, he is not necessary a manipulator. At the very least, it is a borderline case. Perhaps, Nadir just provides a relevant piece of information. Daniel wants to become esteemed, and Nadir tells him about the best way to do it. In this case, the reference to public opinion is relevant and informative, because the content of esteem is derivative of public opinion. The problem is that, from a Kantian perspective, a person in the passionate state is especially disposed towards irrational behavior. A passionate person is, basically, an addict in a broad sense. Daniel does not just desire to be esteemed; he is fully possessed by his desire. If Nadir has his own agenda and knowingly exploits this particularly vulnerable Daniel's state to influence his beliefs and behavior, he is a manipulator.

But what if Daniel is a firm logical egoist and ostentatiously does not care about opinions of others? Well, Nadir may try to present things as if it is, actually, Daniel's own opinion that joining the military is the matter of honour. But while I think that it is possible, I struggle to describe in detail how it can be done. Still, even if Daniel is a firm logical egoist, he does not derive his opinions from nowhere. Logical egoism is dangerous, exactly because it gives a person an unfounded confidence in his independence from the external influence. This influence remains invisible for the logical egoist, because he never tries to be impartial; he never tries to look at things from someone else's position. My point here is that even a logical egoist leans on public opinion, when he forms his understanding of what makes one esteemed. And while an ambitious person may be better protected from direct manipulation than those who are conformable, his active passionate state makes him a much more desired prey for the manipulator.

There are other passions besides ambition, such as greed, mania for power, vengeance, etc. They also can be exploited by a manipulator in a similar fashion. Say, Nadine needs to convince Marcus that if he adopts a certain belief or does a certain thing, it will contribute to realizing his passion. If this act of convincing involves lying, or conscious (on behalf of Nadine) appeals to prejudices, then it should automatically be regarded as manipulation. But if it does not involve them, then, once again, it may be baseless to call Nadine a manipulator, as it is a borderline case. However, if she knows that Marcus is fully possessed by his passion, and uses true statements to exploit Marcus's passion for her own agenda, I believe, we should regard her as a manipulator.

4.3 Enthusiasm and other affects

Affects are similar to passions in that they also preclude rational deliberation, but they do it differently. A passionate man finds it incredibly hard to abandon his passion, but, at least, he can deliberate on it. He can think how to fulfill his passion; moreover, he tends to rationalize it. For instance, an avenger may convince himself that his mania for vengeance, actually, is not a passion, but a moral duty. This rationalization – intended to repress one’s sense of guilt – is what Kant calls self-deception³⁹.

Affect does not allow deliberation about itself and means of its realization. While passion is structurally a desire, affect is rather a very specific feeling⁴⁰; a person in the affected state is unable to compare “this feeling with the sum of all feelings (of pleasure or displeasure)” (Kant, 7:254). Basically, affect disrupts the normal process of choice, and other feelings and corresponding desires just end up not being taken into account. A person does not deliberate on what she is going to do, as an affect almost immediately leads to some action⁴¹ or inactivity. For instance, the affect of anger may lead to an uncontrolled violent behavior, while “the affect of fright [may] produce a scream” (Kant, 25:600). A frightened person does not deliberate on why and how exactly he should scream; it just happens. A manipulator can – by provoking an affective state in a manipulatee – trick her into doing something that is somehow advantageous for the manipulator. There are different methods for different affects: the manipulator can try to infuriate the manipulatee, hoping to bait her into violent behavior. Or, perhaps, the manipulator may try to make his victim incredibly depressed and unable to do anything.

I want to consider one particular affect – enthusiasm. Kant defines it as “the idea of the good with affect” and notes that “it is a stretching of the powers through ideas, which give the mind a momentum that acts far more powerfully and persistently than the impetus given by sensory representations” (Kant, 5:272). So, it is a very intensive feeling, caused by entertaining some moral idea (i.e. the idea of the good), which may lead to a great action. Kant shows an ambivalent attitude towards enthusiasm. On the one hand, in *The Conflict of the Faculties* he regards the enthusiasm of German spectators who follow the developments of the French Revolution “as evidence of a moral predisposition or capacity for morality, a moral tendency” (Clewis, 2009, 3). Kant also says that “it is commonly maintained that without [enthusiasm] nothing great can be accomplished” (Kant, 5:272).

³⁹ Self-deception is a particular example of the secret influence of sensibility on the higher faculty of cognition.

⁴⁰ Perhaps, it is helpful to remind here of Kant’s basic account of human action: “cognition of an object gives rise (sometimes) to a feeling of pleasure or pain, and that feeling gives rise (again, sometimes) to a desire or aversion for the object” (Frierson, 2014, 56).

⁴¹ ‘Action’ should be understood here in a broad sense; e.g. adoption of a new belief is also an action.

However, note, that he does not say that he shares this commonly maintained opinion! Moreover, he believes that “every affect is blind ... for it is that movement of the mind that makes it incapable of engaging in free consideration of principles, in order to determine itself in accordance with them” (Kant, 7:275). It seems that Kant does not regard enthusiasm as something good; he rather implies that its presence signifies that there is something good in human nature, because it shows that people can be moved by the moral ideas. However, the enthusiasm itself is an affect, and as the affect it is not good. It precludes rational deliberation and easily may lead to the very grim consequences. A manipulator may try to provoke enthusiasm in his victim. For instance, Tavish may tell Lycus in the gruesome and exaggerated form (i.e. using framing) that Olivia has done some truly horrendous things, but remained unpunished. Tavish’s story has a close connection with the moral idea of justice. It may cause in Lycus two affects at a time: anger and enthusiasm. As result, Lycus may unreflectively do something that Lycus wants him to do.

4.4 Changing public opinion

Affects can be also used for managing or directing public opinion. Previously, I only discussed how the manipulator can either appeal to public opinion or pretend that he appeals to it. However, if the manipulator has enough resources, he may also try to manage or direct public opinion. Here we are talking about a very specific kind of manipulation – political manipulation. There are three basic ways to do it.

1. The first way is demonstrated in Rousseau’s example of the Court of Honor⁴². Usually, it is a gradual process, as the public opinion needs to be changed imperceptibly. For instance, at first, say, it is inappropriate to be interested in the private life of the ruler, but then step-by-step it may also become inappropriate to discuss his political decisions. These imperceptible changes can be especially efficient in the isolated communities, where it is impossible to consider the external view. However, this gradual change in public opinion can also be caused by the morally neutral means. For example, a modern education may change public opinion on GMOs. I do not think that directing public opinion through education – provided it is free from indoctrination – should be regarded as manipulation from a Kantian perspective, as education strengthens, rather than hinders people’s ability to think for themselves.

⁴² See the section 4.1.

2. The second way to manage public opinion is to just massively falsify it. People largely learn about public opinion through media and sociology. Both of them can transmit false information. If people are conformable, they adjust their attitudes accordingly, and this leads to a real change in public opinion.

3. The third way of directing public opinion involves making use of affects. It allows to drastically change public opinion in a short space of time. As I have said before, affects may lead to rash actions, and ‘action’ here is to be understood broadly, for instance, the adoption of a new belief is also an action. It works as follows, a manipulator transmits information to people in a very specific way, he uses framing to provoke a certain emotion (i.e. affect) in them. For instance, Mr. Petrov is a talking head. He tries to frighten people, exposing (or even lying about) some horrible misdoings, made by the representatives of a certain ethnic group. Mr. Petrov especially focuses that it is made by people belonging to this ethnicity. If he does it alone, he may have some influence on people. But if almost all his colleagues are transmitting the same thing, this influence can be incredibly strong. And what is especially interesting is that afterwards it is quite hard for people to reject this affectedly acquired belief, because they know that it is shared by others. Hence, it is already a matter of conformity. And conformity may cause rationalization of this new belief and, as result, self-deception.

4.5 The attractiveness of conspiracy theories

In the previous chapter I discussed Kant’s account of natural. According to it, purposiveness is “a principle for guiding the investigation” (Kant, 20:236). Something is purposive when it appears as if designed or produced according to the idea of some purpose. A manipulator can misuse teleology for his own ends; he can invent a rather crazy explanatory scheme, i.e. a conspiracy theory. Such a theory may unite a lot of unrelated persons and events into some system; usually it also assumes that there is someone who arranged all these events and who is going to benefit from them.

There are two reasons why the conspiracy theory is a product of the **improper** usage of teleology. First, teleology should be applied only to the real systems. In other words, there should be real connections between different parts of the concerned object. It is reasonable to assume that there are such connections within a living organism. It is also reasonable (but, perhaps, slightly less reasonable) to assume that there are such connections within nature as whole. But those systems that are presented in typical conspiracy theories are usually insufficiently grounded. They either lack any proofs that

there are any real connections between their parts, or they try to justify these connections by appeals to prejudices⁴³. Teleology should not be applied to a non-existing system, it is a fallacy.

Second, Kant explicitly says that we cannot use teleology to determine “whether ... purposiveness is intentional or unintentional” (Kant, 20:236). Hence, even if connections between different persons and events presented in the conspiracy theory are real, teleology cannot prove that there really is a designer of conspiracy who is going to benefit from it. Perhaps, it is just a very complicated system composed of the unlucky contingencies.

Still, people may easily accept a conspiracy theory just because it is natural for them to think teleologically and systematically in order to understand a certain phenomena. A skillful manipulator knows about this tendency and may use it for his own advantage.

4.6 The concept of manipulation

In this section I want to synthesize all my intuitions concerning manipulation in order to elucidate its concept, which can be used for distinguishing between manipulation and seemingly similar, but actually different phenomena.

I have discussed above the various manipulative techniques. All of them involve either lying or bypassing rational deliberation. Rational deliberation can be bypassed either by exploiting a certain psychological state of a manipulatee, or by appealing to the prejudices he has. We can say that manipulation is necessarily epistemically defective. According to Sheryl Tuttle Ross something “is epistemically defective if either it is false, inappropriate, or connected to other beliefs in ways that are inapt, misleading, or unwarranted” (Ross, 2002, 23).

However, there are some phenomena that may be called epistemically defective, but still it would be rather strange to classify them as manipulation. The proprieties are the most obvious example. While good manners are, in a sense, misleading, it would be an overkill to call every well-mannered person a manipulator.

Or, another example, the separation of powers has a certain similarity with manipulation. Different branches are competing with each other, but their ambition serves a different end. They – perhaps, unwillingly – limit each other, thus preventing the concentration of power. As Kant says, even a nation of devils can have good governance, the only thing that is needed is

⁴³ E.g. by appealing to the multitude prejudice: “all believe that A and B are connected”.

to organize them and arrange a constitution for them in such a way that, although they strive against each other in their private intentions, the latter check each other in such a way that the result in their public conduct is just as if they had no such evil intentions (Kant, 8:366).

Should we classify this arrangement as manipulation? Once again, I think we should not. Both examples involve an illusion. The one who is well-mannered creates an illusion that he is also a morally good person. The efficient separation of powers results in an illusion that those who are in power have no evil intentions; moreover, those who are in power may have their own illusions that the goal of this separation is to provide an arena for competition.

But there is a great difference between these illusions and misrepresentations that are characteristic of manipulation. As Frierson notes, “the effect of illusions does not wear off when one knows the truth” (Frierson, 2005, 116). A polite person can still be agreeable to others, even if they know that her politeness does not mean much. A politician usually perfectly knows that his intended ambitious actions have unintended consequences of being check and balances for the actions of others, but it does not make him less ambitious. In other words, these illusions do not necessarily involve misrepresentation or bypassing rational deliberation, they can be efficient even if it is absolutely clear that they are just illusions. It is different in case of manipulation. If one fully acknowledges that he is manipulated into doing something that he does not really want to do, he has a really good motivation, at least, to try to resist manipulation.

The last thing I want to stress here is that a person in question should at least to some extent realize what he is doing in order to be called a manipulator. If Delmar says something to Mia, which suddenly provokes a passionate or affected state in her, it does not necessary mean that Delmar is a manipulator. If Mia’s reaction was impossible to predict, Delmar’s actions cannot be classified as manipulation.

Hence, manipulation is a deliberate action that aims to change the behavior or beliefs through **conscious** lying or **conscious** usage of techniques that preclude or hinder a moral deliberation.

5 Personal autonomy as a solution

In the previous chapters I discussed at length how manipulation is mechanically possible and why it is so efficient. In this chapter I am going to look into how we can protect ourselves from manipulation. In particular, I will consider how Kant understands relations between autonomy and enlightenment. What is of interest to me is whether Kant’s

idea of autonomy can somehow contribute to precluding or resisting manipulation. I will also touch upon the issue of moral responsibility for manipulation.

Earlier, I have argued that manipulation from a Kantian perspective involves either lying or bypassing rational deliberation. Bypassing rational deliberation can have different forms.

- a) When manipulator (A) successfully appeals to the prejudices of the manipulatee (B), the transition between two cognitions in B is carried out by means of unsound connection, resulted “from empirical influence on human cognition” (Frierson, 2014, 196). In other words, B goes from one belief to another using the irrational provisional principle. E.g. A can tell B that everyone believes P, and B may infer from it that she also should believe P. In that case A exploits the prejudice of the prestige of the multitude.
- b) When A provokes an affective state in B, the latter finds herself unable to reflect on her feeling and corresponding desire. E.g. A can tell or show B in the gruesome and exaggerated form that a third person has done some horrendous things. If B lacks strong enough self-mastery, it may provoke an uncontrollable burst of anger in her, bypassing any reflection and leading to the irrational behavior which may be somehow advantageous for A.
- c) When A provokes a passionate state in B, the latter – while she can reflect on how to pursue her passion – finds herself unable to check whether this passion is consistent with morals, happiness as a whole, and with other inclinations. E.g. A can convince B that doing X is a matter of honor⁴⁴. It may cause B to sacrifice her own well-being and high moral standards for a dubious goal. What makes it different from the affect case is that here the manipulatee is not fully precluded from the rational deliberation. For instance, she can be very inventive in her means of pursuing passions.

These cases are idealized. It is possible and even plausible that the most effective cases of manipulation involve simultaneous lying, appeals to prejudices, and attempts to provoke affects and passions in the manipulatee. What unites all these three cases is that the manipulator tries to exploit the manipulatee’s inability or unwillingness to use her own understanding. The latter is the main topic of Kant’s famous essay ‘Answering the Question: What is Enlightenment?’

⁴⁴ Mania for honor is a paradigmatic example of passion. See Kant, 7:268.

5.1 Moral autonomy and personal autonomy

Kant states that the motto of enlightenment is “*Sapere aude*⁴⁵! Have the courage to make use of your own [understanding]!” (Kant, 8:35) If a person lacks the courage, then, according to Kant, she lives in the state of self-incurred immaturity. It is important that for Kant courage (or self-mastery, or moral strength) is the most crucial virtue (see Louden, 2011, 27). If one lacks it one cannot be truly committed to one’s principles. Kant notes that “the very Latin word *virtus* originally signifies nothing else but courage, strength and constancy” (Kant, 27:492). In another essay he restates that “the maxim of always thinking for oneself is enlightenment” (Kant, 8:146n). It seems that the enlightenment maxim is supposed, among other things, to help one defend oneself against manipulation; Kant says that if one adopts this maxim then “superstition and enthusiasm disappear”⁴⁶ (Kant, 8:146n). But what does it mean to think for oneself?

To make use of one’s own reason means no more than to ask oneself, whenever one is supposed to assume something, whether one could find it feasible to make the ground or the rule on which one assumes it into a universal principle for the use of reason (Kant, 8:146n).

This enlightenment maxim shares some similarity with Kant’s formula of the universal law: “Act only in accordance with that maxim through which you can at the same time will that it become a universal law” (Kant, 4:421). This similarity becomes even more noticeable in the third critique, where Kant notes that this maxim is “of a reason that is never passive. The tendency toward the latter, hence toward heteronomy of reason, is called prejudice” (Kant, 5:294). What is important here is that Kant contrasts the activity of the enlightened reason with its heteronomy⁴⁷. It implies that the goal of enlightenment is the achievement of autonomy. It is striking that in the essay on enlightenment Kant does not mention autonomy at all. In the following I am going to investigate whether being autonomous helps one to protect oneself from being manipulated.

Autonomy – as Kant defines it – is the capacity of the rational will to be self-legislative (see Kant, 4:432-433). There are two aspects of exercising autonomy. The first aspect is negative; according to it the autonomous will is not governed by any external cause. Human nature is the external cause for Kant, hence any action caused exclusively by

⁴⁵ Latin phrase meaning ‘Dare to know’.

⁴⁶ According to Kant, superstition is “the greatest prejudice of all” (Kant, 5:294) and enthusiasm is “the idea of the good with affect” (Kant, 5:272).

⁴⁷ Passions and affects hinder rational deliberation; hence they also lead to heteronomy.

an inclination or instinct cannot be considered autonomous⁴⁸. That is why actions caused by passions and affects are not autonomous. The second aspect of exercising autonomy is positive. According to it, autonomy does not amount to arbitrariness, i.e. it still presumes very specific laws. And while these laws are self-prescribed, they are, at the same time, universally true for every rational being. These laws are the laws of morals, hence Kant's autonomy often referred in the secondary literature as moral autonomy. It is coherent with what Kant says: "[a]utonomy of the will is the sole principle of all moral laws and of the duties conforming to them" (Kant, 5:33).

Kant treats moral autonomy as a capacity (see Kant 4:438; O'Neill, 1990, 75-76; Formoza, 2013, 194-195). All people who have a capacity to follow the moral law are autonomous, that is why they are responsible for their choices. They do not need to exercise autonomy in order to be regarded as autonomous, as the possible legislators of moral law. It means that, for Kant, almost all people are morally autonomous, only those who are severely mentally impaired can be excluded. But it is definitely not the case that only severely mentally impaired people are susceptible to manipulation. However, there is a very quick solution to this: it is possible to distinguish between being autonomous as having a capacity of autonomy and being autonomous as exercising this capacity. Actually, Kant himself makes the distinction between the will (*Wille*) and choice (*Willkür*) (see Kant, 5:36). While every person has the autonomous will, it does not mean that she always chooses autonomously (see Formoza, 2013, 195).

In a sense, Kant's idea of autonomy gives us infallible knowledge. More precisely, it gives us an infallible way to check whether a certain maxim (a practical principle) is morally permissible; we just need to check whether it is universalizable⁴⁹. But it seems that the applicability of the universalizability test for protecting one from manipulation is quite limited. In some cases it may be obvious that a certain maxim imposed on a person by a manipulator is immoral per se. For instance, the manipulator may persuade the manipulatee that someone should be killed. But while the maxim of killing someone is obviously incompatible with the moral law, the manipulatee is unable to properly conduct the universalizability test if she is in the affective or passionate state. Hence, there is a circle: you cannot check the maxim if you do not exercise autonomy in the first place.

⁴⁸ It does not mean that any influence of psychology automatically leads to heteronomy. If a person is fully determined by a rational principle to do something, and her inclinations (virtues) assist her in doing it, her action is autonomous. Or, conversely, even if a person is fully determined by a rational principle to do something, and her inclinations hinder her from doing it, she still can be successfully moved by her principle; and in that case her action is autonomous too.

⁴⁹ If the maxim generates a contradiction when universalized, then it is not compatible with the moral law.

Moreover, when Kant talks about the lack of courage to use one's own understanding, he does not talk not only about the purely moral sphere. It seems that manipulation may involve communicating maxims that are not directly incompatible with the moral law, but still very undesirable for the manipulatee. For example, someone can be manipulated into support of the law, which is going to be very disadvantageous for him in the long term. My initial intuition here is that in this case the manipulatee does not adopt the morally dubious maxim⁵⁰, but still he is obviously a victim of manipulation.

Kant's idea of autonomy is closely connected with morals. It may seem that we need the concept of personal rather than moral autonomy to deal with manipulation. Personal autonomy involves "governing [oneself] in the pursuit of [one's] own conception of the good" (Formoza, 2013, 193). According to it, "the free choice of goals and relations" is "an essential ingredient of individual well-being" (Raz, 1988, 369). Manipulation compromises one's ability to choose what is right or good in one's own opinion. In other words, manipulation deprives one of personal autonomy. I argue that the enlightenment maxim endorses personal autonomy, as it can be applied to the variety of cases, some of which are not related to the morals. But it does not follow from this that personal autonomy does not contribute to morals from a Kantian perspective. I hold that there are, at least, two important conceptual connections between moral autonomy and the enlightenment maxim:

1. The enlightenment maxim necessarily follows from the demands of morals. The refusal to think for oneself is a failure to respect humanity in one's own person. In the enlightenment essay Kant clearly says that it is the duty "of every human being to think for himself" (Kant, 8:36). And from that perspective there was something morally dubious about the manipulatee's maxims if he did not adopt the enlightenment maxim when he was manipulated into support of the undesirable law. As Schneewind points out, "[t]he failure of courage to use our own reason is a moral failure" (Schneewind, 2010, 316). Kant even explains the cause of this moral failure. It lies in radical evil of human nature. Kant's understanding of evil is rather unusual:

Whether the human being is good or evil, must not lie in the difference between the incentives [of self-love and morality] that he incorporates into his maxim ... but in their subordination (in the form of the maxim): which of the two he makes the condition of the other. It follows that the human being (even the best) is evil only because he reverses the moral order of his incentives in incorporating them into his maxims (Kant, 6:36).

⁵⁰ He may truly believe that this law is very beneficial not only for him, but for the whole humankind. Endorsing such the law is not only compatible with the demands of morals; endorsing it is an imperfect duty to others.

The human beings have a propensity to evil. In other words, they have propensity “knowingly and consistently to choose maxims contrary to the moral law” (Louden, 2011, 33). As I argued, the enlightenment maxim follows from the moral law. When people consciously choose maxims which are based on the inclinations (e.g. on laziness), they break the moral law.

It should be noted, however, that someone who has adopted the enlightenment maxim can still be fooled by a manipulator. In that case, he does nothing wrong. What I want to stress here is that the enlightenment maxim is not the guaranteed defender against manipulation. As I will show in the next chapter, it allows checking the provisional principles on which one bases one’s judgments, but not the judgments themselves. However, by checking the principles we may improve our chances on making our judgments more reliable. Thus, the enlightenment maxim is beneficial for theoretical reason. In a sense it provides a connection between theoretical and practical reason, with the former subordinated to the latter (see O’Neil, 1990, 24-27).

2. The second connection between autonomy and the enlightenment maxim can be found within Kant’s philosophy of history:

Here the first true steps are taken from brutishness to culture, which consists, actually, in the social worth of human beings. And here all of the talents are gradually developed, taste is formed, and, even, through continual enlightenment, the beginning of a foundation is laid for a manner of thinking which is able, over time, to transform the primitive natural predisposition for moral discernment into definite practical principles and, in this way, to ultimately transform an agreement to society that initially had been pathologically coerced into a moral whole. (Kant, *Idea*, 8:21)

Enlightenment is important, because through it human beings can become moral. According to Kant, nature itself cannot make people moral, as morals is the matter of freedom; however, nature can by means of antagonism coerce human beings into the culture⁵¹ which is understood as “[t]he production of the aptitude of a rational being for any ends in general (thus those of his freedom)” (Kant, 5:431). The enlightenment maxim necessarily follows from the demands of morals and, at the same time, governs theoretical reason, it helps us to be more successful at making judgments and, hence at attaining our ends. Thus, the enlightenment contributes to what Kant calls the culture of skill which can be defined “as the capacity to attain the ends that humankind sets for itself, whatever they may be” (Allison, 2009, 24). Eventually, it may lead to the establishment of cosmopolitan civil society and the perpetual peace, which is the precondition for the full development of all human predispositions (see Kant, 8:25), including that to personality. “The

⁵¹ I discuss it in details in the third chapter.

predisposition to personality is the susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice” (Kant, 6:27). Hence, the perpetual peace is a precondition for moralization of humanity⁵².

5.2 Who is to blame?

I am going to end this chapter with a more detailed discussion of the moral responsibility for manipulation. It is easy to see why a manipulatee can be regarded as prudentially responsible for her own misfortunes: she was not prepared and vigilant enough, so she fell a victim to manipulation. But it seems that, from a Kantian perspective, a manipulatee should also be regarded as morally guilty for not respecting her own humanity. Some may see it as a quite undesired consequence of a Kantian theory of manipulation. Is it a good theory, if it puts blame on a victim? Well, I have a few things to say about it.

First, the moral responsibility for not respecting one’s own humanity is not necessary related to manipulation. On the one hand, if a person does not think for herself, she is already guilty for not respecting her own humanity. She does not need to be manipulated in order to be morally responsible for not thinking for herself. On the other hand, even if a manipulator fails at influencing his victim’s beliefs and actions, i.e. even if there is no actual victim, a manipulator is still to blame, as he does not respect humanity in the face of the other (i.e. a potential manipulatee). Therefore, these are two distinct responsibilities, which are not really connected. The guilt of the manipulator does not depend on whether he is successful or not. The guilt of a person who is unwilling to think for herself also does not depend on whether she ends up being manipulated or not.

Second, while the responsibility for manipulation is fully on the manipulator, we can regard a manipulatee as prudentially and morally responsible for being weak in the face of manipulation. Perhaps, it sounds not very inspiring, but this kind of responsibility does not justify manipulation in any sense. Instead, a Kantian theory of manipulation shows that we have a potential that can be realized. In other words, it stresses the fact that every person can (and ought to) contribute to her own security from manipulation.

Third, even that responsibility for being weak in the face of manipulation is not homogeneous. Kant primarily speaks about people who are **unwilling** to think for themselves. He also compares this unwillingness with immaturity. It means that he does

⁵² However, it cannot force people to become moral (see Kant 7:93), as the moral progress is the result of spontaneous acts of freedom. From a Kantian perspective, the moral progress cannot be ‘caused’ by any external force.

not expect, for instance, that kids should be morally responsible for being manipulated. The same can be said about people who are deprived of any possibility to develop their understanding. Moreover, Kant says that those people who are in the affected state should not be considered responsible neither for their state, nor for the actions they do in this state⁵³. An affect is “a lack of virtue and, as it were, something childish and weak, which can indeed coexist with the best will” (Kant 6:408). It completely bypasses any rational deliberation. According to Kant, a person cannot be held accountable for an action, which was fully automatic. Of course, a person has a duty to cultivate his virtues, but Kant defines this duty as wide or imperfect. “Wide duties are duties to set ends. We are required to have these ends, but it is left up to us what and how much we do in pursuit of them” (Wood, 1999, 44). Hence, technically, even if a person is doing almost nothing in order to cultivate her virtues, she cannot be regarded as doing something morally bad.

Fourth, a person is not obliged to do something impossible. Kant claims that ‘ought’ always presupposes ‘can’ (cf. Kant, 3:A548/B576). If a person does everything she can and still ends up being manipulated, it would be wrong to regard her as prudentially and morally responsible for this. While she has a certain potential for defending herself from manipulation, this potential is socially dependent. One may do as much as it is possible in these particular social conditions. For example, one cannot properly think in a broad-minded way, if the public use of reason (e.g. free speech) is severely limited.

6 What can we do?

In the previous chapter I assumed that the enlightenment maxim to always think for oneself may contribute to our defense against manipulation by establishing personal autonomy. I also argued that the enlightenment maxim necessary follows from the demands of morals. We need to use our own understanding in all spheres, because, otherwise, we fail to obey the moral law in that that we fail to treat humanity with respect in our own person. In this chapter I want to consider in details what it means – to use one’s own understanding – and how precisely it can help us to resist manipulation. I will argue that it involves three interconnected thought experiments that are akin to the universalizability test of moral autonomy. I will also argue that these three thought experiments are actually three different perspectives on the one and the same thought experiment.

⁵³ I am talking here only about moral responsibility. From a Kantian point of view, it does not necessary follow from this that people in the affected state are also released from legal responsibility.

In some his texts Kant adds to the enlightenment maxim – **to think for oneself** – two others that are closely connected to it: **to think in the position of everyone else** and **always to think in accord with oneself**. In the *Lectures on Logic* Kant terms these three maxims “[u]niversal rules and conditions for avoiding error” (Kant, 9:57), in the third critique – “the maxims of the common human understanding” (Kant, 5:294), in the *Anthropology from a Pragmatic of View* – “[t]he precept for reaching [wisdom]”, where wisdom “is an idea of a practical use of reason that conforms perfectly with the law” (Kant 7:200). It should be stressed that all these maxims are supposed to govern not our judgments themselves, but the principles on which we base our judgments, i.e. provisional principles⁵⁴. In the following I will examine these maxims one by one and will argue that they are three different formulations of one principle⁵⁵.

6.1 The maxim of reason's self-preservation

The first maxim is the maxim of enlightenment – to think for oneself. Kant also calls it the maxim of reason's self-preservation. Here is how it works:

To make use of one's own reason means no more than to ask oneself, whenever one is supposed to assume something, whether one could find it feasible to make the ground or the rule on which one assumes it into a universal principle for the use of reason. This test is one that everyone can apply to himself; and with this examination he will see superstition and enthusiasm disappear, even if he falls far short of having the information to refute them on objective grounds (Kant, 8:146n).

Kant basically proposes a test for provisional principles. As Onora O'Neill nicely puts it, the provisional principles should be “neither self-stultifying nor self-defeating in use” (O'Neill, 1990, 38). If our provisional principle fails this test by justifying a false judgment, then it should be discarded. In order to check the principle we need to try to imagine any possible situation where the judgment based on this principle turns out to be false. If such a situation is impossible, then this principle is a proper principle of understanding (in a broader sense).

However, this test seems to be too demanding. I will illustrate it with an example: suppose that Yuval has made nine true statements, and we know it for sure. Does our experience of previous Yuval's truthfulness justify trust in his future statements? According to Kant, it does not. Because it is not unthinkable that the person who had been truthful nine times would lie in the future.

⁵⁴ See the second chapter for the detailed account of provisional principles.

⁵⁵ It is akin to three different formulations of the moral law in the *Groundwork of the Metaphysics of Morals*.

Kant discards any appeal to any kind of a reputation as a proper principle of justification. It is a safe position, but it seems to be counter-productive, because very often we just have to base our judgments on the non-ideal principles. What makes this position even stranger is the fact that Kant regards induction as a proper principle of understanding⁵⁶ and calls it the source of comparative universality (see Kant, 3:B124/A91), whereas it is quite possible to imagine a situation where reliance on induction leads to the false judgment. What is the difference between relying on induction in the theoretical matters and relying on reputation in the practical matters? As I see it, for Kant, the practical matters are fundamentally different, because they involve rational beings characterized by transcendental freedom⁵⁷. The only authority that everyone ought to recognize is the authority of reason; the authority of reputation never can be universal.

It may be argued that while the good reputation of the source cannot be the reason for taking the information from it for granted, it may be prudent to take the good reputation of the source as a reason for paying attention to what it communicates. But, of course, you still need to use the proper principles of understanding (i.e. not the appeal to the reputation), when you are assessing this information. This pragmatic principle may help one to deal with the information overload, as Kant acknowledges that prejudices can be “true provisional judgments; what is wrong is only that they hold for us as ... determining judgments⁵⁸” (Kant, 9:75-76). Still, it seems that this pragmatic principle is not compatible with the strong epistemic requirements of the enlightenment maxim, as it is quite easy to imagine a situation where this principle disables one from getting the information from the truthful source with the bad reputation.

As I have already stated in the previous chapter, the enlightenment maxim endorses the cultivation of the virtue of courage: “Have the **courage** to make use of your own [understanding]!” (Kant, 8:35) Kant treats courage (or self-mastery, or moral strength) as the most crucial virtue (see Loudon, 2011, 27), if one lacks it one cannot be truly committed to one’s principles. Courage or self-mastery helps one to cope with affects and

⁵⁶ More exactly, a principle of the power of judgment.

⁵⁷ Perhaps, it would become clearer if we apply to the issue the concept of possible worlds. According to Kant, it is possible to imagine that this particular person could have done otherwise, even under the virtually same circumstances; but it is absolutely impossible to imagine that the natural laws governing the nature could be different, as, for Kant, they are necessary (i.e. the same in the every possible world). You can apply induction to inanimate and non-rational entities, because they are governed exclusively by natural laws. In contrast with them, the rational beings as noumenons are also subject to the causality of freedom; hence, the appeal to the reputation cannot be a reliable principle. Of course, Kant himself does not use the concept of possible worlds. I use it here only for the illustrative purpose.

⁵⁸ The difference is that a prejudice may lead to the right judgment. Hence, it may have some provisional merit. However, it would be wrong to think that the rightness of the judgment follows from the prejudice. In other words, the prejudice cannot justify the right judgment.

passions, hence it can be said that it is the most important virtue one should have in order to defend oneself from manipulation.

The enlightenment maxim also helps us to be more successful at making judgments and, hence at attaining our ends. Thus, it contributes to what Kant calls the culture of skill which can be defined “as the capacity to attain the ends that humankind sets for itself, whatever they may be” (Allison, 2009, 24). Kant perceives some passions⁵⁹ as the product of the social development. According to him, they can be pacified by the further socio-cultural development⁶⁰. Thus, by contributing through the enlightenment to the cultural development one makes passions less dangerous.

6.2 The maxim of a broad-minded way of thinking

According to the second maxim, one should think in the position of everyone else. One is thinking in a broad way

if he sets himself apart from the subjective private conditions of the judgment, within which so many others are as if bracketed, and reflects on his own judgment from a universal standpoint (which he can only determine by putting himself into the standpoint of others) (Kant, 5:294n).

The second maxim stresses the pluralism of reason, but it also has connection with Kant’s idea of universality, which was already introduced in the previous maxim. Already in the first critique Kant stresses that reason is necessarily public, “and its pronouncement [in people] is never more than the agreement of free citizens, each of whom must be able, without holding back, to utter his qualms indeed, even his veto” (Kant, 3:A739/B767). Why is the pluralism of reason important? Every human being has propensity to subordinate the laws of reason to her own inclinations (i.e. propensity to evil). Very often it is necessary to look at one’s own situation from the impartial position in order to understand that the principles on which you justify your judgments are wrong, as you may overlook some incompatibilities between standpoints of different people which would arise were these people to base their judgment on the provisional principle in question. Hence, in addition to the enlightenment maxim we need another test or thought experiment, which is aimed to help the one to judge from a universal standpoint⁶¹.

⁵⁹ For instance, Kant holds that envy, ingratitude, spite, manias for honor, power, and wealth have a social origin.

⁶⁰ I will elaborate on this in the third and fourth chapters.

⁶¹ Cf. with Hume’s concept of the extensive sympathy and the general point of view (see Hume, 1978, 581).

However, the efficiency of this test depends on the possibility of free communication. One cannot even imagine oneself speaking from the authority of reason, if there is no possibility for the public discussion. The best way to find the proper principles of understanding “is to encourage the increasingly public use of reason” (O’Neill, 1990, 38). The public use of reason is associated with scholar’s right to discuss and criticize all deficiencies in any sphere. According to Kant, every civilian can be such scholar in her free time (see Kant, 8:36). All of laws and regulations are subject to discuss and change. Any generation should have an opportunity to call in question any decision made before it. The problem is that the public use of reason is actually something which is not under our personal control. However, as I already showed, the adoption of the enlightenment maxim may contribute to socio-cultural development and may eventually lead to the establishment of cosmopolitan civil society and the perpetual peace (see Kant, 8:25). It is possible to argue that it is our wide duty to strive for freedom to make the public use of one’s reason possible in all matters. But in the end it is not something that you can achieve purely on your own. It is more like a long-term strategy to limit external threats to reliability of our judgments.

Public use of reason has another interesting consequence which may be of use for coping with manipulation. One’s attempt to judge from a universal standpoint helps one to cultivate the virtue of courage or self-mastery. In order to get to the universal standpoint, she must be willing to listen to the others; it is not an easy goal and involves an exercise of the virtue of self-mastery. Moreover, the universal standpoint may have an effect on our inclinations. For someone, who lives in the isolated community, the demands of propriety can be devastating. It may be inappropriate to discuss the king here, or to criticize the laws. The sense of propriety may make a person a weak-willed conformist in this community⁶². However, if one is a part of the cosmopolitan community and judges from the universal standpoint, then the effects of propriety may be quite beneficent both for morals and reliability of one’s judgments.

Kant believes that propriety “gave the first hint toward the formative education of the human being as a moral creature” (Kant, 8:112). Hence promoting propriety should be regarded as a wide duty towards oneself and others (see Kant, 6:473). The attempt to be polite reveals to oneself, and helps to develop, self-mastery. Through self-mastery one is able to resist one’s self-deception about the impossibility of virtue. In other words, self-mastery helps one to prevent self-exemption from the moral law based on the false belief

⁶² This negative effect of propriety was discussed in the fourth chapter.

that one is just unable to be virtuous. And I have already discussed the role of self-mastery in coping with manipulation.

6.3 The maxim of consistency

According to the third maxim, one must always think in accord with oneself. Kant says that it is “the most difficult to achieve, and can only be achieved through the combination of the first two and after frequent observance of them has made them automatic” (Kant, 5:295). It is the basic regulative principle of rational thought and, at the same time, it implies our duty to strive for perfection of our reason. Our hierarchy of maxims should be consistent. The maxims of the moral law should have the highest priority. Others should be combined in the coherent manner as well. For example, if a person knowingly chooses something that is moderately useful in the short-term perspective and extremely disadvantageous in the long term perspective, then he obviously bases his judgments on the unsound provisional principles.

It is important to stress that second and third maxim work as check-and-balances system for the first one. The unreflective adherence to the first maxim alone easily can lead person to the state of self-assurance which can even prevent him from realizing that his rational will is determined by external causes. I think there is a striking similarity between these three maxims. It seems that it is just impossible to adopt one of them without adopting others as well. If one wants to think for oneself, she needs to find the universal principle which is correct for everyone and does not produce any inconsistencies. If one wants to think in the position of everyone else, she needs to be impartial, it is impossible without thinking for oneself; she also wants to find the universal principles; it is impossible without thinking consistently. And if one wants to think consistently, she needs to find the universal principles which are correct for everyone, once again, it is impossible without thinking for oneself. Therefore, I argue, that these three principles are in fact the three different formulations of one and the same principle. Adopting it involves a very sophisticated thought experiment which helps us to check our provisional maxims, to cultivate self-mastery and, eventually, to contribute to the socio-cultural development of the humankind.

It is a normative principle; it tells us how we should think; it does not tell how we really think. For Kant, ‘ought’ presupposes ‘can’ (see Kant, 3:A548/B576), and it seems that following this principle is not logically impossible. On the other hand, it is not logically impossible to be saint, however no one succeeds. Kant’s anthropology shows that

he was fully aware that the human beings are not perfect and they live in an imperfect society, and all its imperfection is the result of their free choices. Still, he shows that it is no logically impossible that we can change it for the better. In an imperfect society adopting the enlightenment maxim may be counter-productive; sometimes, it is more productive to be guided by prejudices here. But at the very least, the enlightenment principle leads us to recognize that prejudices, passions and affects may lead us up a blind alley, and that there is always another possibility.

Conclusion

In my thesis, I have argued that Kant's pragmatic anthropology and empirical psychology can explain us why people are susceptible to manipulation. I have shown that people's susceptibility to manipulation can be causally explained by some defects in cognition and volition, namely by prejudices, affects and passions. I have argued that the manipulative techniques are efficient because they are essentially interwoven with the underlying principles of social order.

A Kantian theory of manipulation has good explanatory power and provides us with an unorthodox perspective on a manipulatee. It considers a manipulatee as prudentially and morally responsible for being vulnerable in the face of manipulation. This kind of responsibility does not justify manipulation in any way, but underlines that every person can contribute to her own security from manipulation. I have shown that it can be done through striving for personal autonomy. Personal autonomy is to be established through adoption of Kant's three maxims of common understanding. I have argued that these three maxims are, in fact, three different formulations of one and the same maxim. Its adoption involves a very sophisticated thought experiment which helps us partially protect ourselves from manipulation.

Abstract

In my thesis, I argue (1) that Kant's pragmatic anthropology and empirical psychology can provide us with an explanation why people are susceptible to manipulation. I show that people's susceptibility to manipulation can be explained by some defects in cognition and volition, namely prejudices, passions and affects. From a Kantian perspective, the manipulative techniques are efficient because they are essentially interwoven with the underlying principles of social order. I also argue (2) that Kant's

maxims of common understanding can help us to partially protect ourselves from the deteriorating effects of manipulation by establishing personal autonomy. I show that the adoption of the maxims of common understanding involves a very sophisticated thought experiment which helps us fight with prejudices and calm our affects and passions. My project can be summed up as the reconstruction of Kant's theory of manipulation.

References

Citations from the Critique of Pure Reason are located by reference to the pagination of Kant's first 'A' and/or second 'B' editions. All other passages from Kant's works are cited by the volume and page number, given by Arabic numerals separated by a colon, in the Academy edition of Kant's writings.

Allison, H. (1990). *Kant's Theory of Freedom*, Cambridge University Press, New York.

Allison, H. (2009). 'Teleology and history in Kant: the critical foundations of Kant's philosophy of history' in Rorty, A.O. (ed.) and Schmidt, J. (ed.) *Kant's Idea for a Universal History with a Cosmopolitan Aim. A Critical Guide*. Cambridge University Press, New York.

Clewis, R. (2009). *The Kantian Sublime and the Revelation of Freedom*. Cambridge University Press, New York.

Formoza, P. (2013). Kant's Conception of Personal Autonomy, *Journal of Social Philosophy*, 44(3): 193-212.

Frierson, P.R. (2005). The Moral Importance of Politeness in Kant's Anthropology, in *Kantian Review*, Vol.9, March, 2005, 105-127, DOI: 10.1017/S1369415400002028.

Frierson, R. P. (2014). *Kant's Empirical Psychology*. Cambridge University Press, New York.

Hume, D., Selby-Bigge, L.A. (ed.) (1978). *A Treatise of Human Nature*, 2nd ed. Oxford: Clarendon Press.

Kant, I. (1991). *The Metaphysics of Morals*. Cambridge University Press, New York.

Kant, I. (1996a). *Critique of Pure Reason*. Hackett Publishing Company, Inc., Indianapolis.

Kant, I. (1996b). *Religion and Rational Theology*. Cambridge University Press, New York.

Kant, I. (1999). *Correspondence*. Cambridge University Press, New York.

Kant, I. (2000). *Critique of the Power of Judgment*. Cambridge University Press, New York.

- Kant, I. (2002a). *Groundwork for the Metaphysics of Morals*. Vail-Ballou Press, New York.
- Kant, I. (2002b). *Critique of Practical Reason*. Hackett Publishing Company, Inc., Indianapolis.
- Kant, I. (2006a). *Anthropology from a Pragmatic Point of View*. Cambridge University Press, New York.
- Kant, I. (2006b). *Toward Perpetual Peace and Other Writings on Politics, Peace, and History*. Vail-Ballou Press, New York.
- Kant, I. (2012). *Lectures on Anthropology*. Cambridge University Press, New York.
- Louden, R.B. (2000). *Impure Ethics. From Rational Beings to Human Beings*. Oxford University Press, New York.
- Louden, R.B. (2011). *Kant's Human Being. Essays on his theory of human nature*. Oxford University Press, New York.
- O'Neill, O. (1990). *Constructions of reason. Explorations of Kant's practical philosophy*. Cambridge University Press, New York.
- Raz, J. (1988). *The Morality of Freedom*. Oxford University Press, New York.
- Rousseau, J.-J. (1987). 'Discourse on the Arts and Sciences' in *Basic Political Writings of Jean-Jacques Rousseau*. Hackett Publishing Company, Inc., Indianapolis.
- Rousseau, J.-J. (2004). *Letter to D'Alembert and Writings for the Theater*. Lebanon, NH: Dartmouth College published by University Press of New England.
- Schmidt, A. (2016). Sources of evil or seeds of the good? Rousseau and Kant on needs, the arts and the sciences, in Lifschitz, A. (ed), *Engaging with Rousseau: Reaction and Interpretation from the Eighteenth Century to the Present*, Cambridge: Cambridge University Press, pp. 33–56.
- Schneewind, J.B. (2010). *Essays on the History of Moral Philosophy*. Oxford University Press, New York.
- Taylor, R. (2005). 'Kantian Personal Autonomy' in *Political Theory*, Vol. 33, No. 5 (Oct., 2005), 602-628.
- Wood, A. (1999). *Kant's Ethical Thought*. Cambridge University Press, New York.
- Wood, A. (2007). *Kantian Ethics*. Cambridge University Press, New York.

Non-exclusive licence to reproduce thesis and make thesis public

I, Semen Reshenin,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:
 - 1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and
 - 1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,
- ‘Kant on manipulation and personal autonomy’, supervised by Eva Piirimäe.
2. I am aware of the fact that the author retains these rights.
3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **18.05.2017**

Semen Reshenin.